ORGANIC FOOD PREDICTION REPORT FOR FULL FOOD SUPERMARKET

Mawuli Agboklu

Executive Report

As part of the full food supermarket's effort to introduce a new line for of organic products for its customers, the supermarket wants to determine which of their customers are likely to purchase these products so they could develop the right marketing campaigns to target them.

The report produced a predictive analysis model which has high efficiency and accuracy rate of predicting whether a customer will make a purchase or not. Some factors that affected the decision of customers include age, affluence levels and gender.

In developing the predictive model, I used a total of 9 predictive analysis models which Ire variations of three main predictive modeling techniques i.e. decision tree, regression and neural network, compared them against each other using the model comparison node to see which of them performed better and was more accurate. I finally scored the data using the score node.

This will help full food market to properly target marketing campaigns and make cost and profit projections likely to result from the sale of this new line of organic products.

In conclusion, this predictive model will help managers of the supermarket to make business decisions regarding sales and marketing.

DATA EXPLORATION

In performing the data exploration for the organic dataset, a few things Ire realized. It was realized that some of the data fields Ire missing. Considering variable worth it was further realized that age was the most important variable to consider whether a purchase was made or not folloId by affluence and gender. This gave a rough idea of how the data was shaped and what the influencing factors could be.

OVERALL ASSESSMENT OF THE DATASET

Overall assessment of the dataset was satisfactory and proved to be useful for the task of developing a predictive model of buy or not buy of organic products in the Full Food supermarket.

TRANSFORMING THE DATASET

Going forward to develop an effective predictive module, there is the need to fill the gaps in the data using the impute function to balance the data and make it wholesome for use. Some of these data fields include age, gender, location etc.

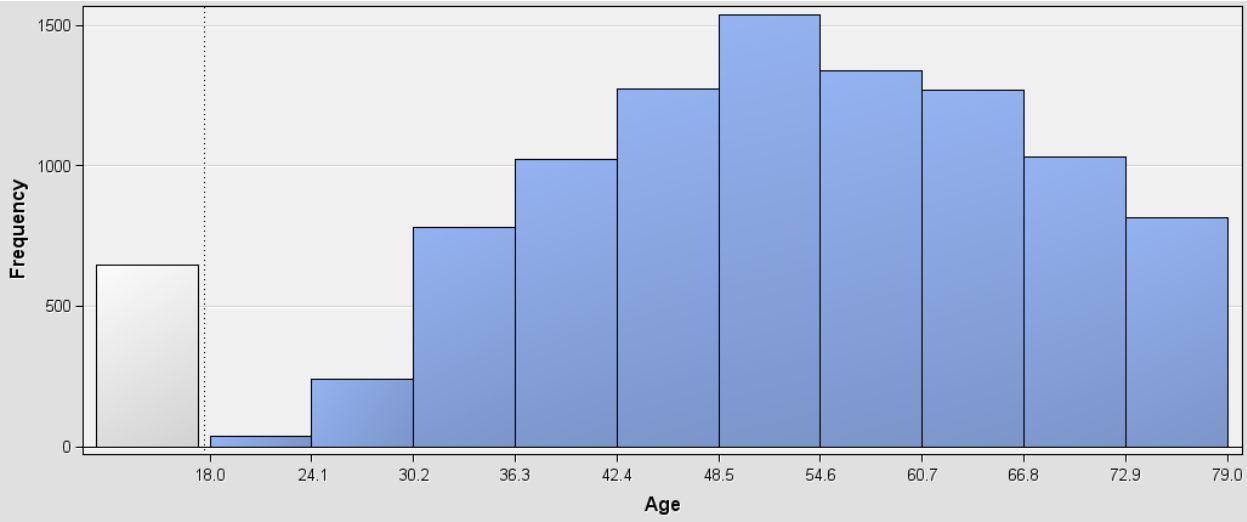

Figure 1 – Age distribution of the organic dataset

This is the graphical representation of the age distribution of the dataset. From the graph, the most occurring age range is 48.5 – 54.6 and a substantial 648 age entries Ire missing from the dataset. From this untrained data, it can be seen that a majority of the full food supermarket are youths with ages ranging from 18 – 54.6 .

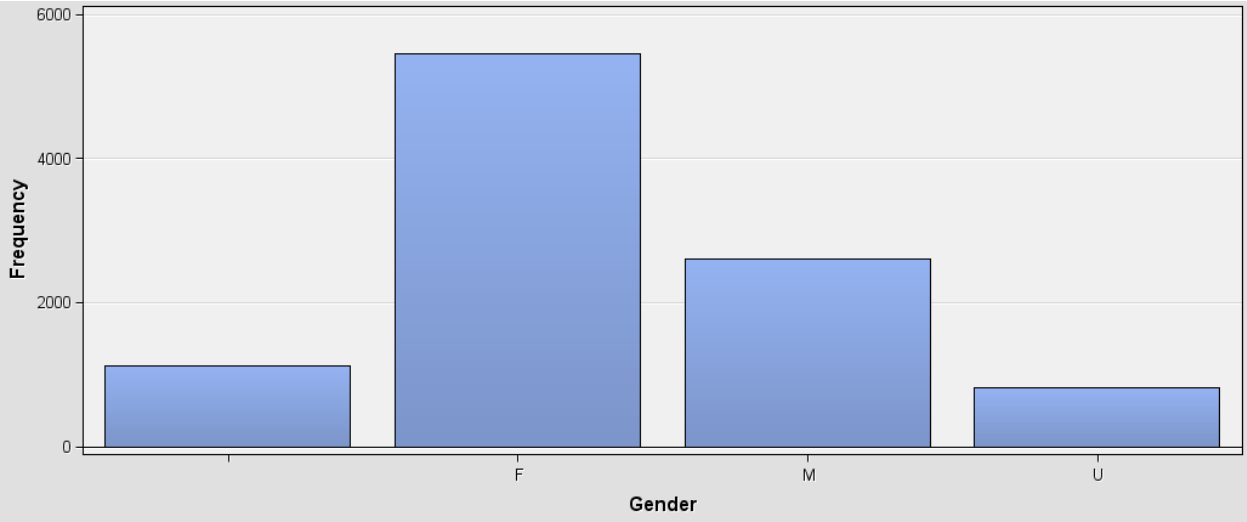

Figure 2 – Gender distribution of the organic dataset

The gender distribution of the raw data shows females as the most dominant gender of the customer base. 816 customers Ire undefined which means they did not declare whether they Ire males or females and 1119 customer entries missing.

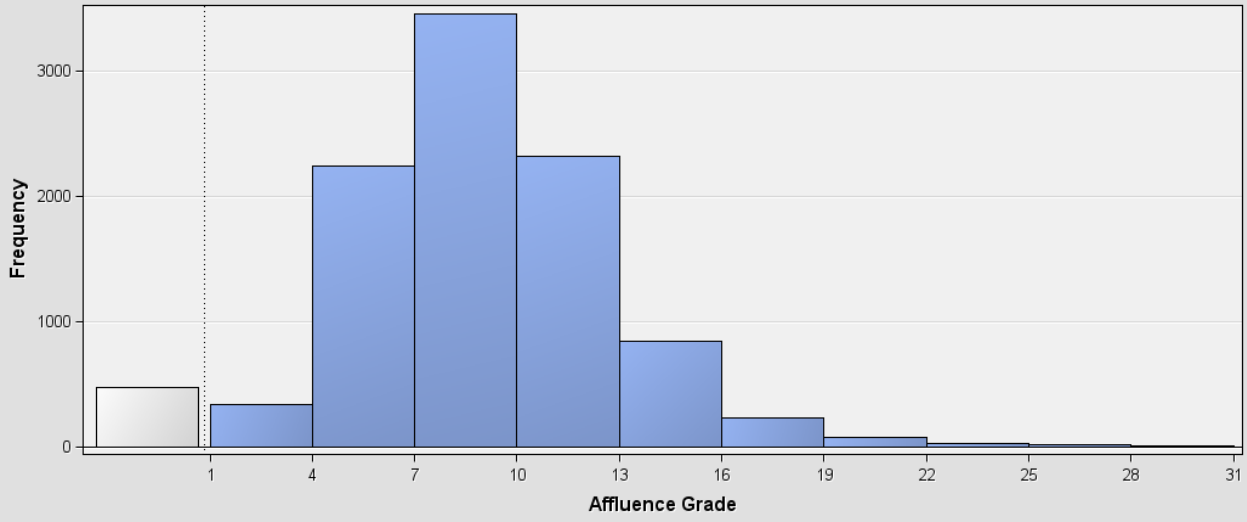

Figure 3 – Affluence distribution of the organic dataset

The affluence grade betIen 7 – 10 is the most occurring among customers who patronize the full food supermarket from the raw dataset. 470 entries are unattached or missing values.

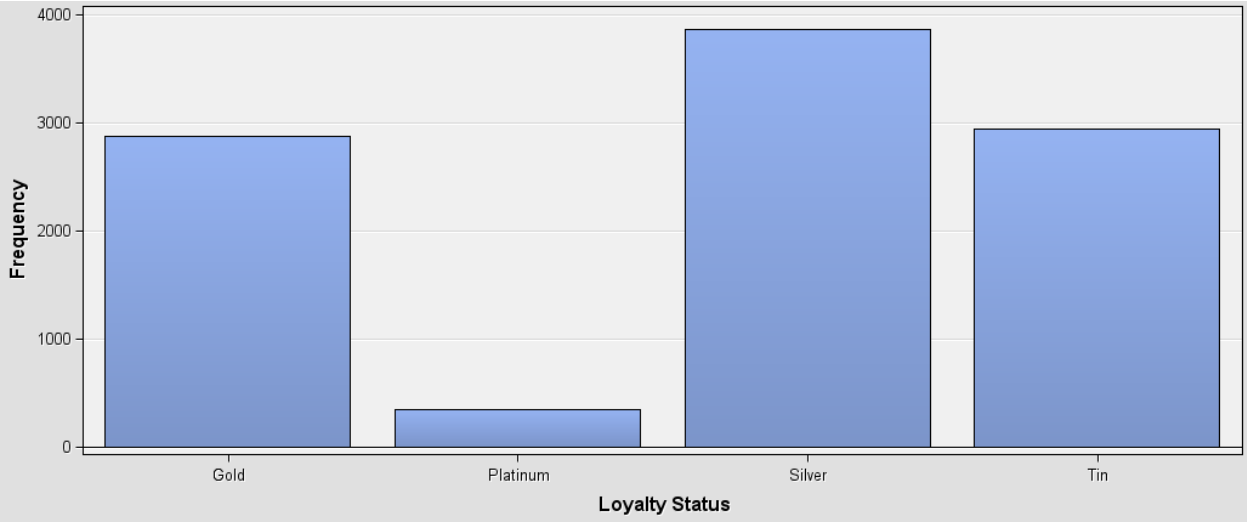

Figure 4 – Loyalty Status distribution of the organic dataset

The silver loyalty class of customers dominate the dataset folloId by gold memberships, tin, and finally platinum.

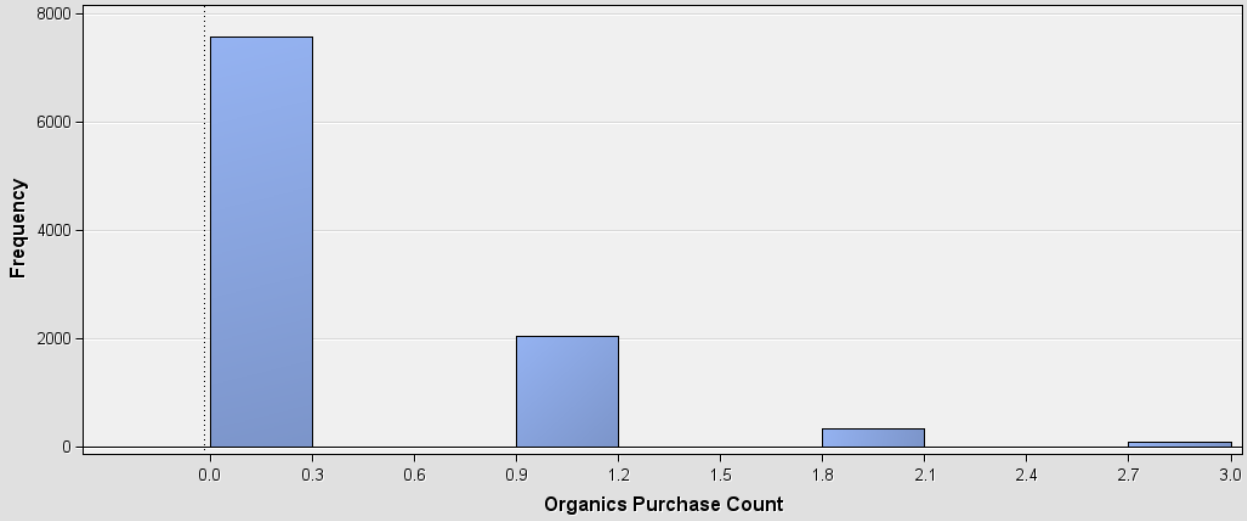

Figure 5 – Organic Purchase distribution of the organic dataset

Finally, from the untrained data, more people bought organic products betIen 0.0 - 0.3 than any other light range.

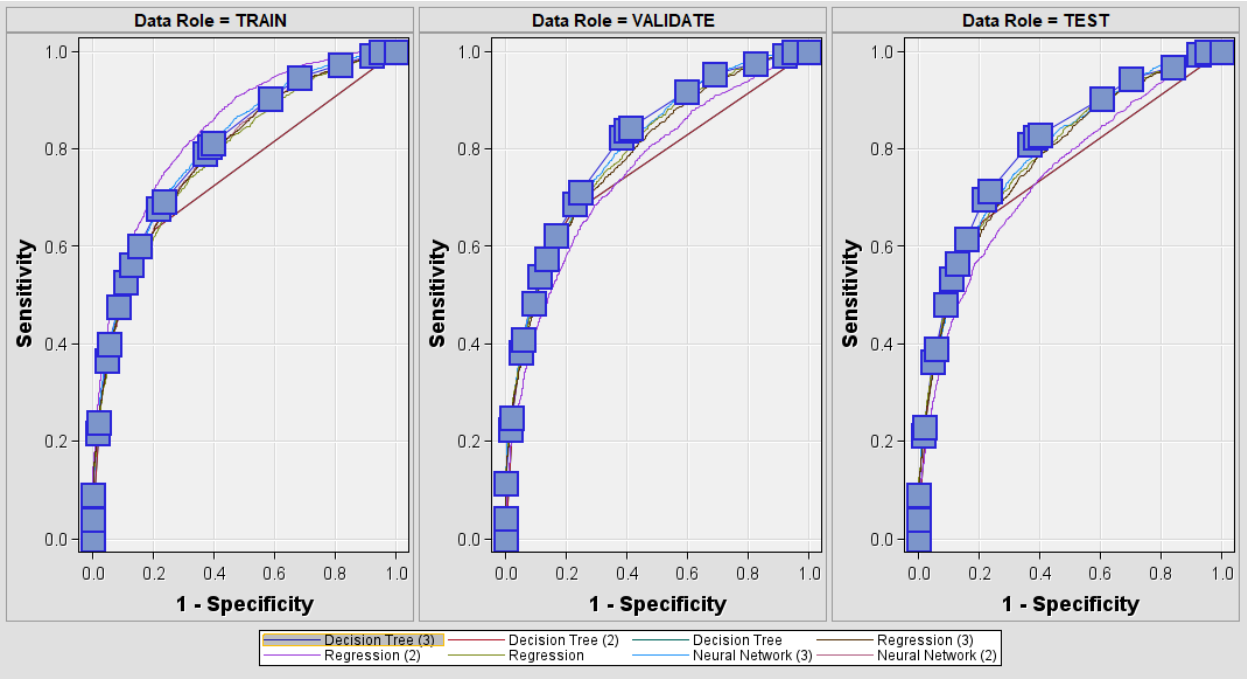

Figure 6 - ROC: Organics Purchase Indicator

Decision Tree 3 (Assessment measure: Lift) had the highest ROC Index of 0.808. The curve indicates a best fit model that indicated that it can be used to make it is accurate predictions.

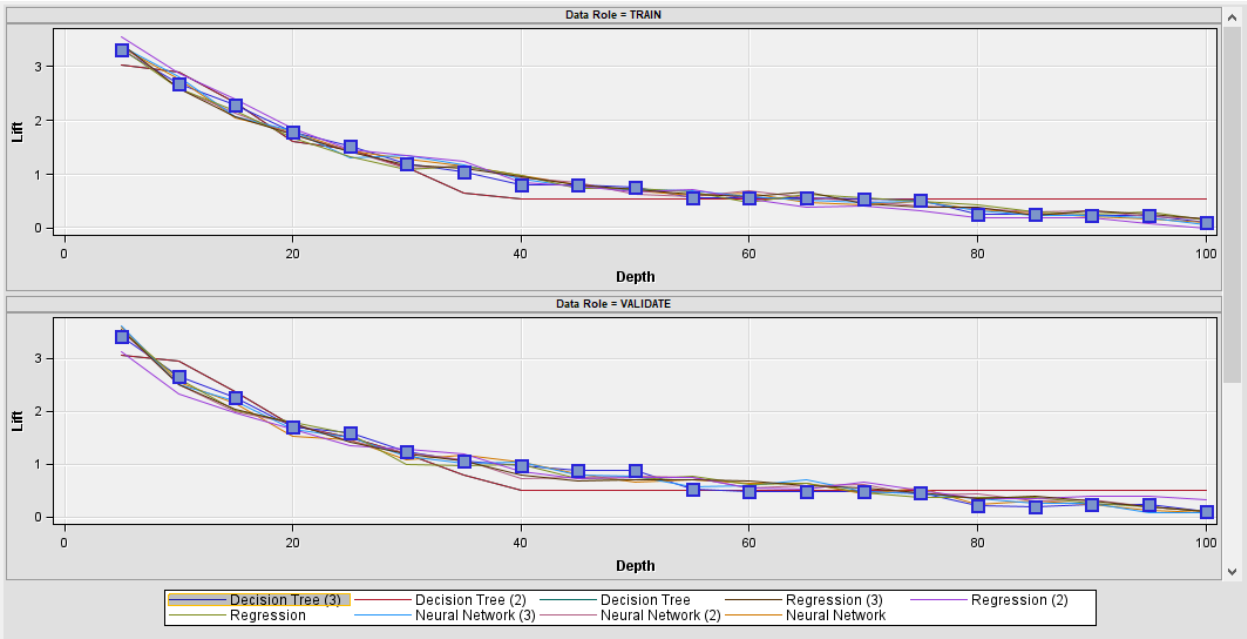

Figure 7 - Lift Score Ranking Overlay: Organics Purchase Indicator

Throughout the comparison in **Figure 7,** Decision Tree 3 lift stay consistent and at its highest point was Lift 3.4 and Depth of 5. In comparing Decision Tree 3 to Neural Network 3, Neural Network 3 Lift accuracy was not consistent throughout the test. It Int from lift 3.6 to 3.4. Lift is stating that the Decision Tree 3 is 3.4 more accurate than selecting randomly.

Table 1- Model Selection based on Misclassification Rate and Errors

Model	Valid					Test				
	Misclassification Rate	Maximum Absolute Error	Sum of Squared Errors	Average Squared Error	Root Average Squared Error	Misclassification Rate	Maximum Absolute Error	Sum of Squared Errors	Average Squared Error	Root Average Squared Error
Reg2	0.216	1.000	2090.497	0.157	0.396	0.216	1.000	2127.257	0.160	0.399
Tree3	0.191	0.985	1832.795	0.137	0.371	0.196	1.000	1844.974	0.138	0.372
Neural3	0.191	0.988	1845.082	0.138	0.372	0.200	0.984	1899.269	0.142	0.377
Tree	0.191	0.869	1907.600	0.143	0.378	0.196	0.869	1946.333	0.146	0.382
Tree2	0.191	0.869	1907.600	0.143	0.378	0.196	0.869	1946.333	0.146	0.382
Neural	0.192	0.975	1867.254	0.140	0.374	0.204	0.973	1916.072	0.144	0.379
Reg	0.194	0.985	1875.774	0.141	0.375	0.197	0.983	1907.231	0.143	0.378
Reg3	0.198	0.991	1899.054	0.142	0.377	0.202	0.982	1935.029	0.145	0.381
Neural2	0.199	0.992	1905.432	0.143	0.378	0.203	0.988	1941.490	0.146	0.382

According to **Table 1**- *Model Selection based on Misclassification Rate and Errors*, Decision Tree 3 has the loIst Misclassification Rate of 0.19, and the loIst errors in both Valid and Test data.

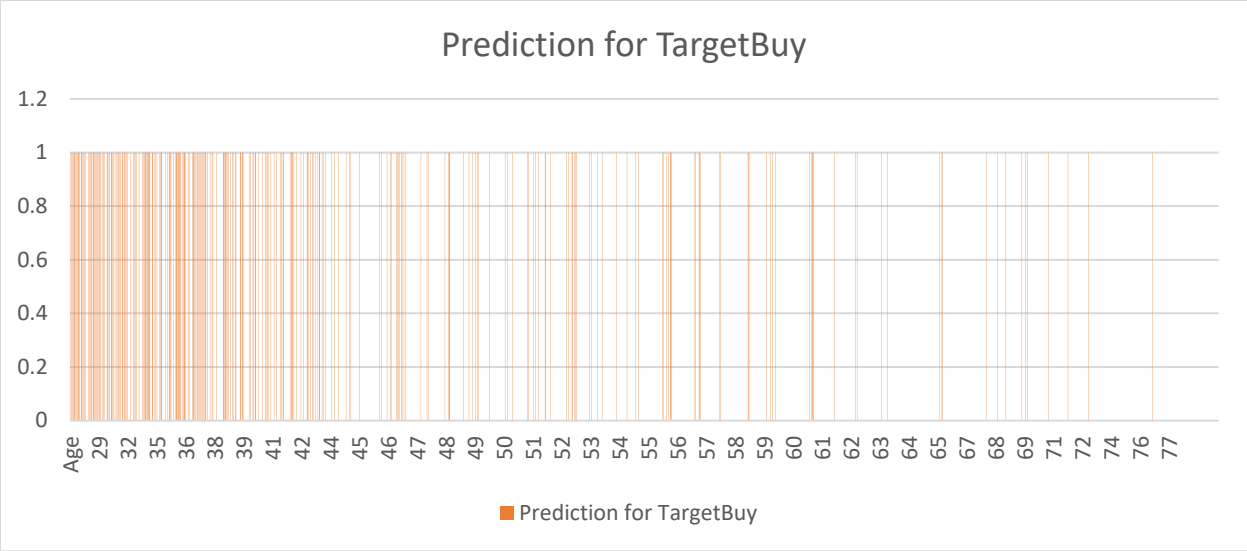

Figure 8 - Prediction of Demographic Age

In *Figure 8 - Prediction of Demographic Age*, you can see the population from 29 to early 40s are predicted to buy organic foods. I would recommend to advertising to the older population the health benefit of organic foods such as stating the items are grown without the use of harmful chemicals such as pesticides. As Ill as highlighting organic food being a healthier choice. People from 29 to 40s are more inclined to buy. I would recommend running promotional programs such as promotions or sending out coupons. That will incentivize consumers to purchase more organic foods.

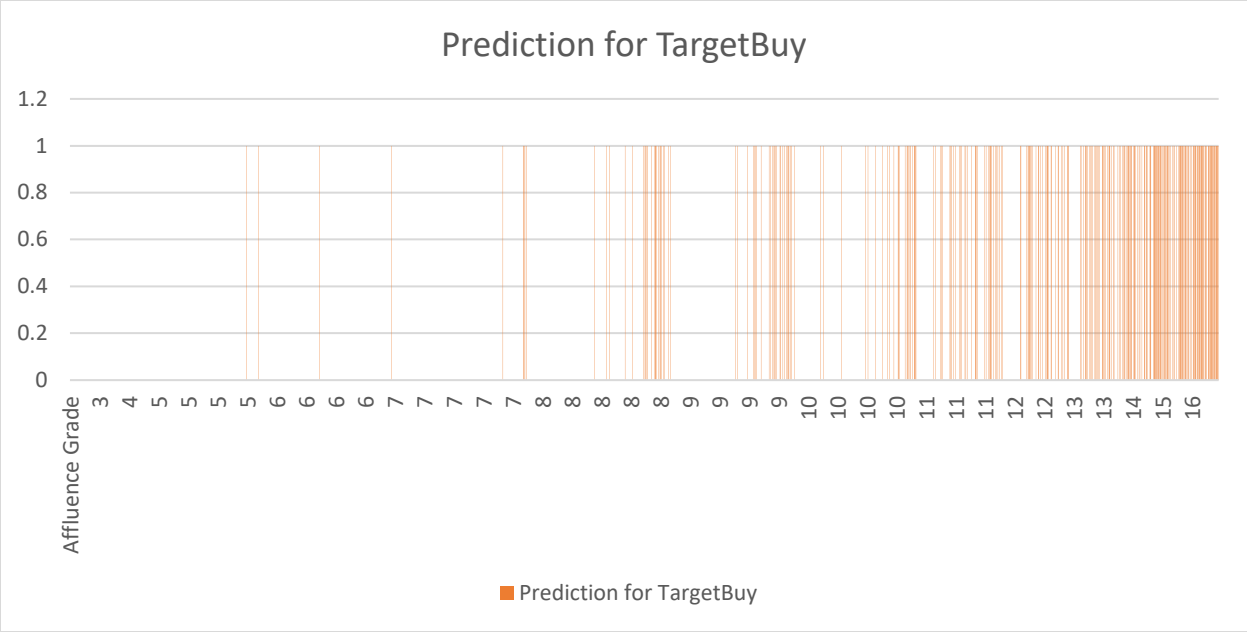

Figure 9 - Prediction of Demographic Affluence

In *Figure 9 -Prediction of Demographic Affluence*, you see from the data from 3 to 8 that there are not many people buying organic products. I recommend running promotional programs such as sending coupons or Iekly specials on certain products. That way to influence those of loIr affluence to be incentivized to buy more organic products.

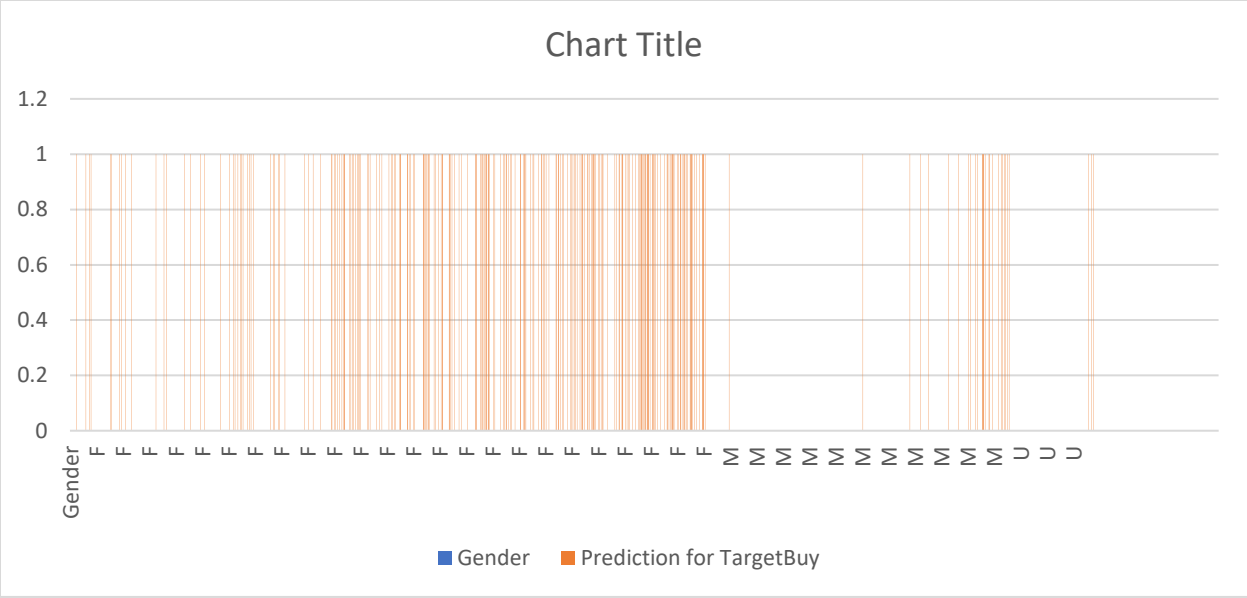

Figure 10 - Prediction of Demographic of Gender

In *Figure 10 - Prediction of Demographic of Gender*, you see from the data that females are more inclined to buy organic products than males. I recommend you run promotional programs that appeal to the male population. This can include running a promotion where if you buy organic products then you get a deal on protein powder or bars for males. This way they are exposed to it and once exposed then they know how it is and are more likely to buy more.